Bellabeat data analysis case study

Gabriel Scozzarro

21/4/2021

1.0 Introduction

Bellabeat is a high-tech manufacturer of health-focused products for women, with a high the potential to become a larger player in the global **smartdevice** market.

By 2016, Bellabeat had opened offices around the world and launched multiple products. Bellabeat products became available through a growing number of online retailers in addition to their own e-commerce channel on their **website**. The company has invested in traditional advertising media, such as radio, out-of-home billboards, print, and television, but focuses on digital marketing extensively. Bellabeat invests year-round in Google Search, maintaining active Facebook and Instagram pages, and consistently engages consumers on Twitter. Additionally, Bellabeat runs video ads on Youtube and display ads on the Google Display Network to support campaigns around key marketing dates.

The products and services offered by Bellabeat are:

- **Bellabeat app**: The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.
- **Leaf**: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.
- **Time**: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.
- **Spring**: This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.
- **Bellabeat membership**: Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness based on their lifestyle and goals.

1.1 Scope of work & Business task

Urška Sršen, the actual Chief Creative Officer in Bellabeat, strongly believe that available consumer data would reveal more opportunities for growth. She request the marketing analytics team to focus on a Bellabeat product and analyze smart device usage data in order to gain insight into how people are already using their smart devices. Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.

In addition, she ask to perform the analysis on data taken from non-Bellabeat smart devices and apply the obtained insights on one Bellabeat product.

The information requested are the following:

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

The expected output are the following:

You will produce a report with the following deliverables:

1. A clear summary of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of your analysis
5. Supporting visualizations and key findings
6. Your top high-level content recommendations based on your analysis

2.0 Data

The data set provided was the **FitBit Fitness Tracker Data** realeased on **Kaggle**, under the license **CC:0 Pubblic Domain**, by the user **Mobius**.

This Kaggle data set contains personal fitness tracker from 33 fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring with a time span of 31 days starting from 04/12/2016. It includes information about daily activity, steps, and heart rate that can be used to explore users' habits.

This data set has a very limited time span and no information on the selection of the user sample were provided by the author. No anagraphic data of the users were provided to double check the sampling, we are forced to assume that the population sample were made correctly.

The data set was copied and stored on my local machine for the analysis. Since the data set is public no security issue should occur.

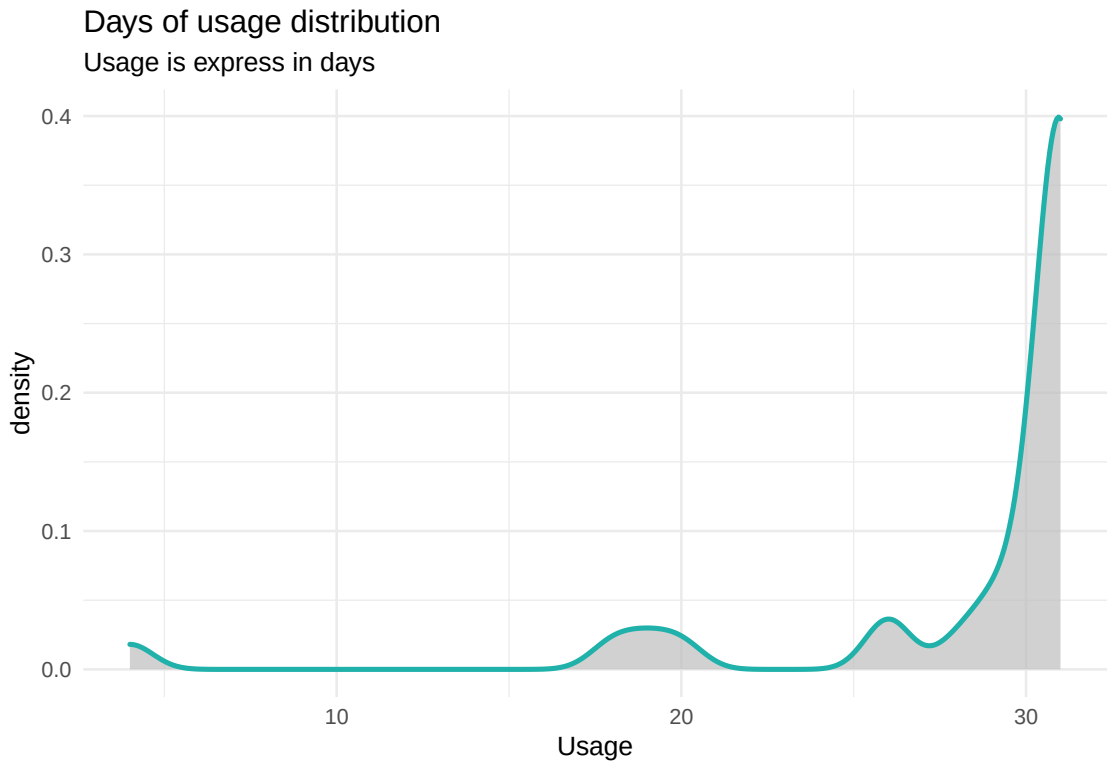
3.0 Tools and process

The analysis was performed using R coding language. A complete list of R packages, a data log and the source code are all avaible on the project github repository at this **link**.

4.0 Analysis

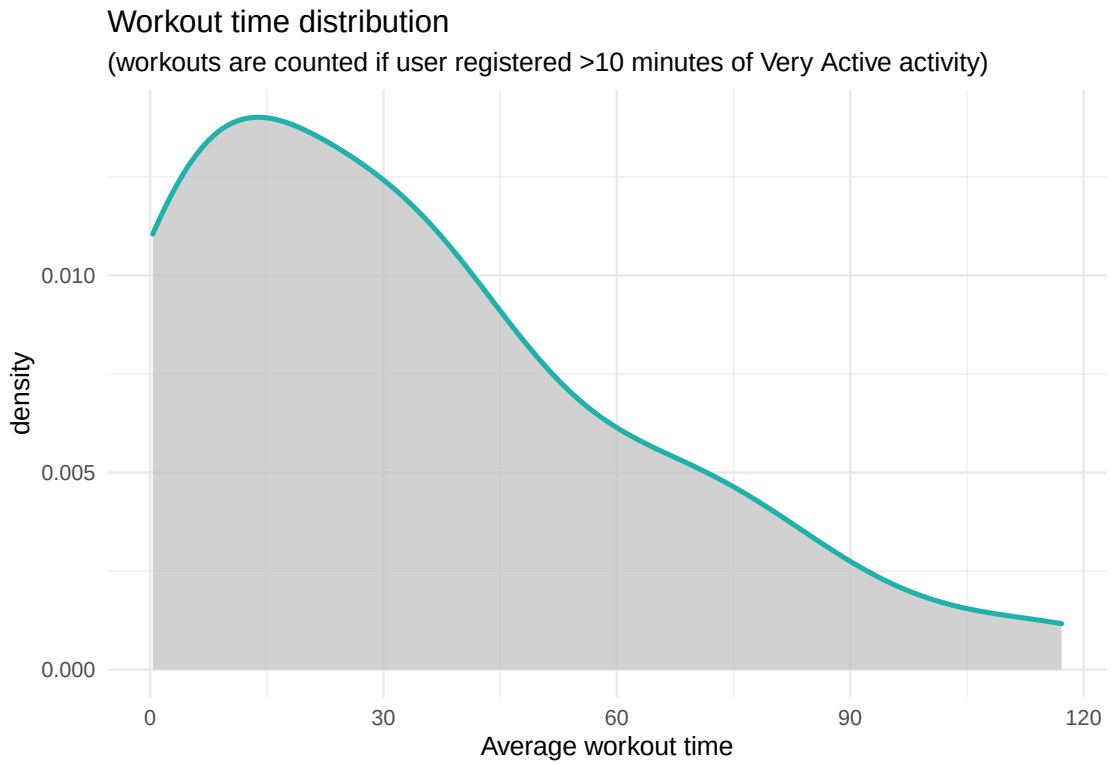
4.1 Daily activity analysis

The daily activity data located inside the Fitbit data set was submitted by all the 33 users. According to this data we discover that among the 33 users the large majority of them use the smart activity for almost all the time period:

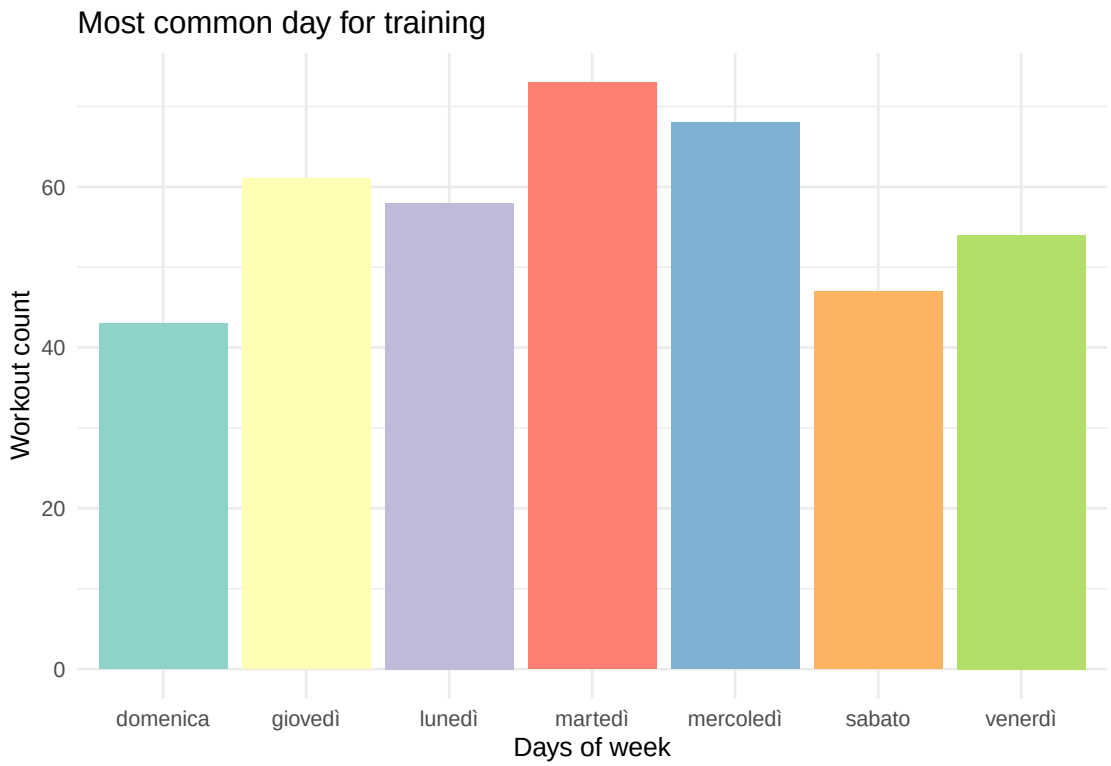


During this time period users performed several workouts with an average of 3 workouts per week.

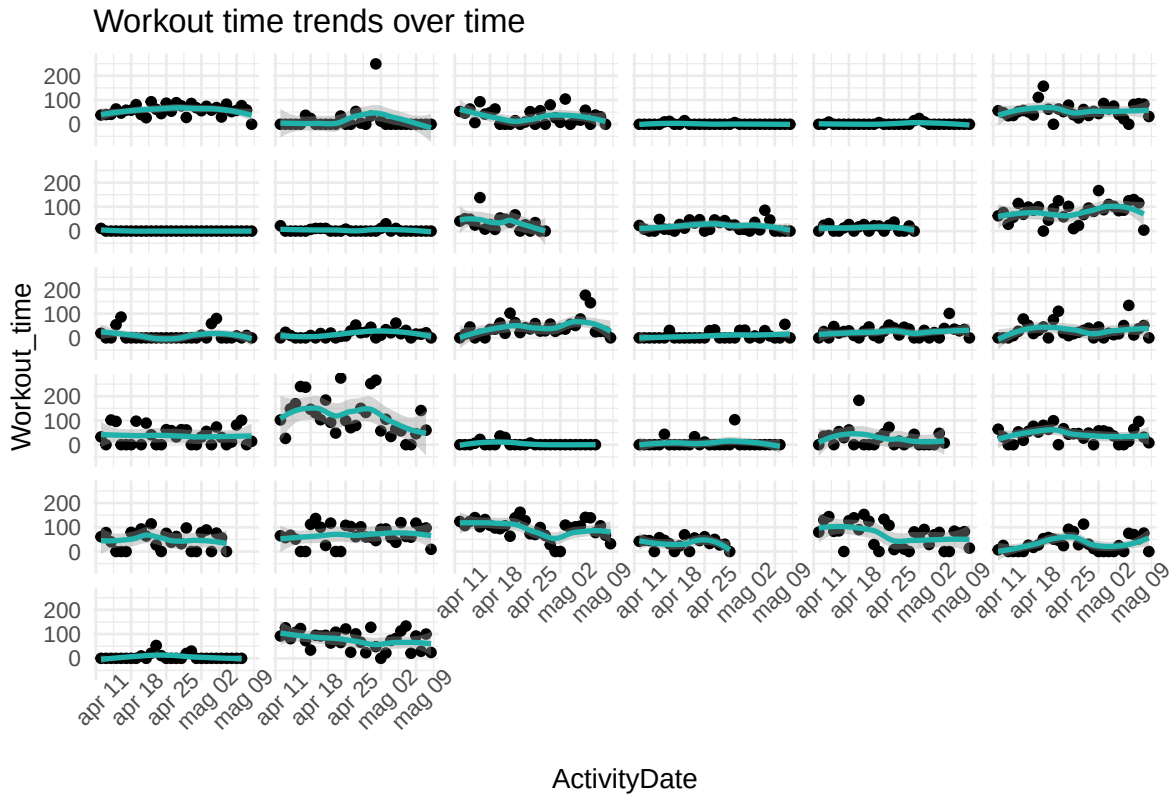
The quantity of activity for each week was 100.7070263 minutes per week which, according to **World Healthcare Organization**, are below the recommended level of 150 minutes per week.

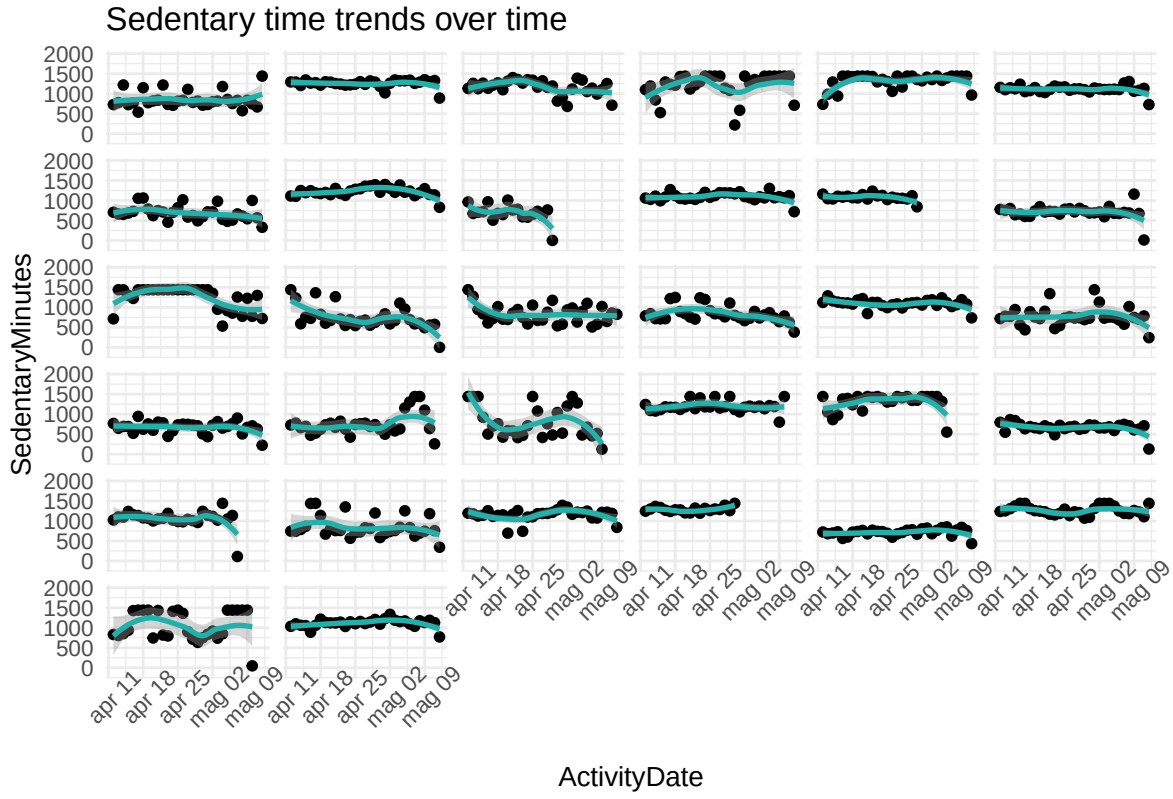


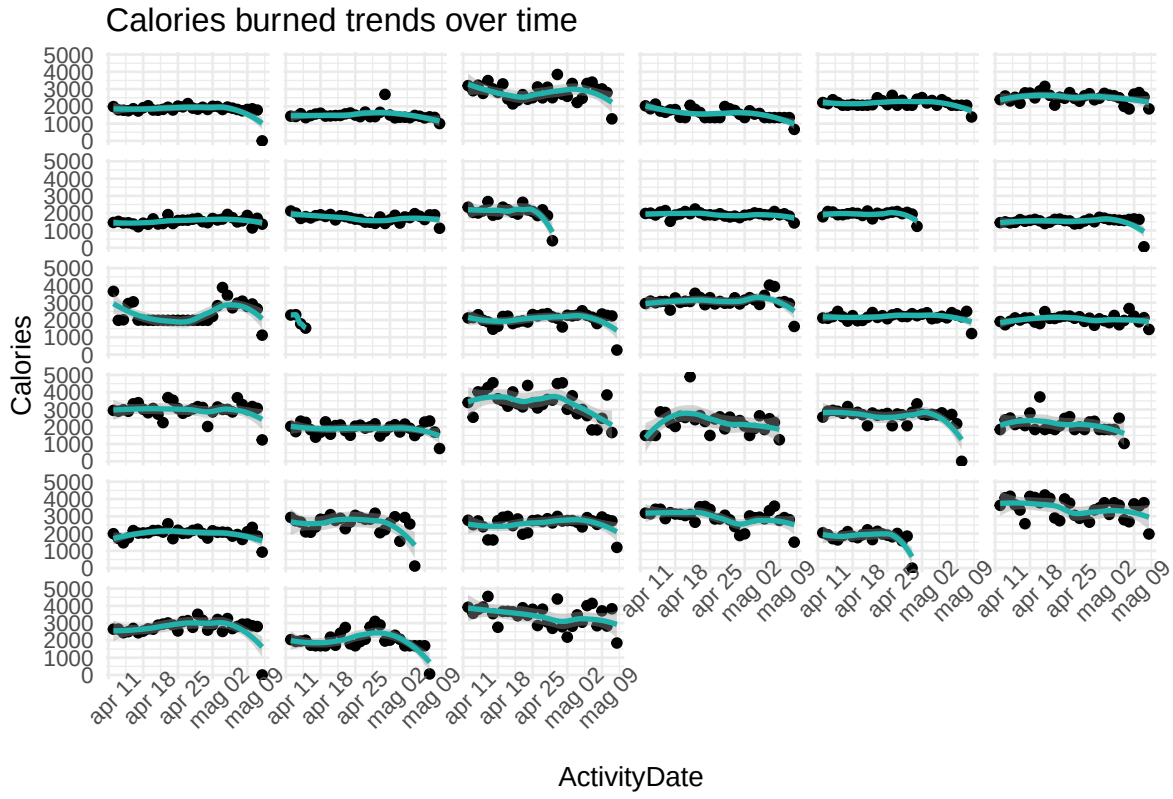
The favorite day for workout was identified on Tuesday:



Since physical exercise is an incremental activity every progress and trend can be observed over time.

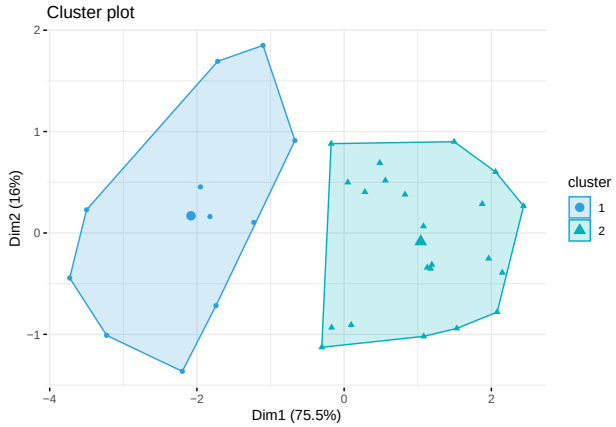


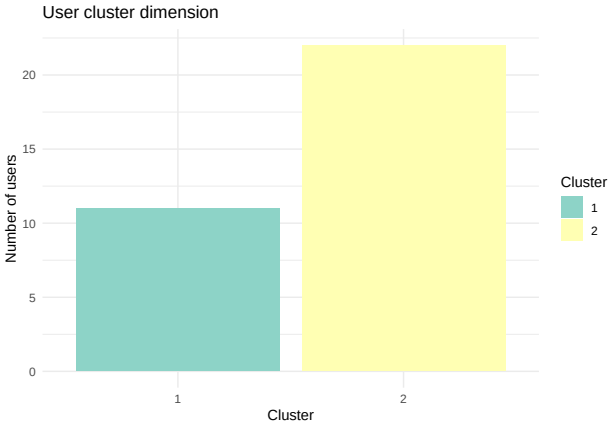


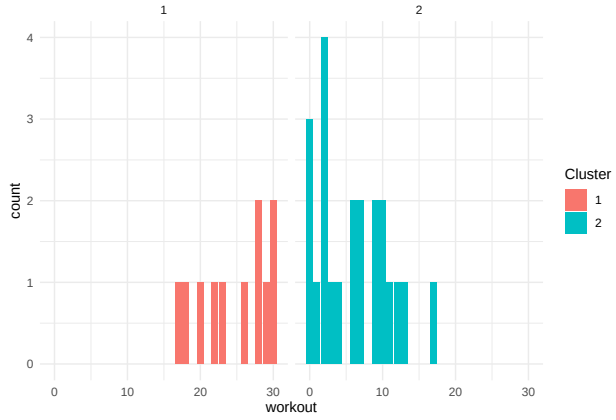


We can observe that over all there no significance changes over time on the workout habit for the user. They don't train more and don't workout harder which is express by the calories burned over time plot. That said, it is important to point out that this data set is very limited in time span and physical activity habits are very difficult to change in this short time.

From the data we can appreciate a pattern based on the workout habits and intensities. Using a **K-means clustering** method we identify two cluster of users:

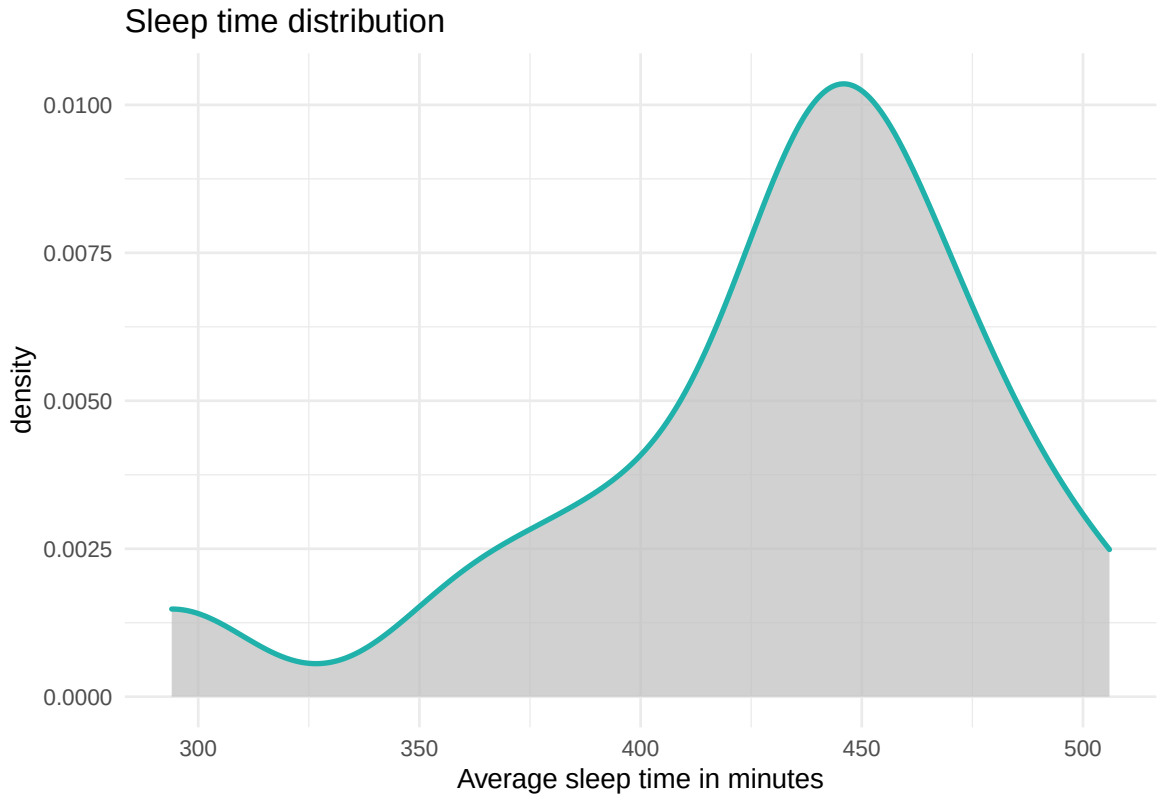




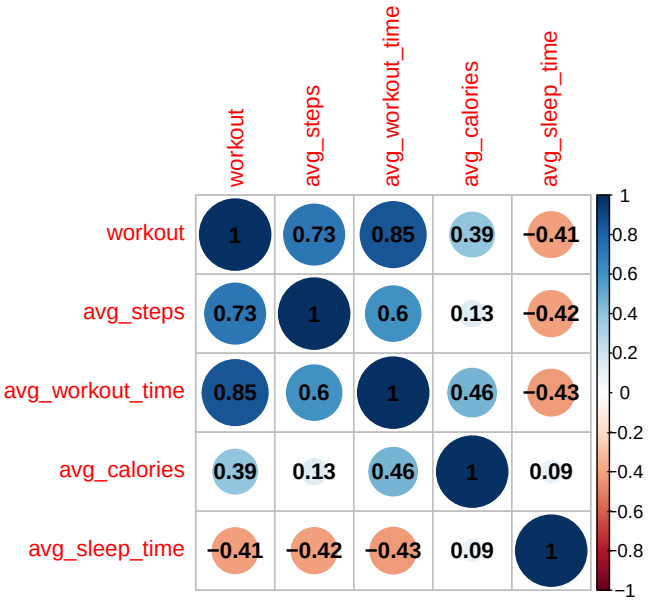


4.2 Sleep analysis

The data set contained also some data on sleep time. The average sleep time is 7 hours which is the minimum level suggested by the all scientists including the **Mayo Clinic**.

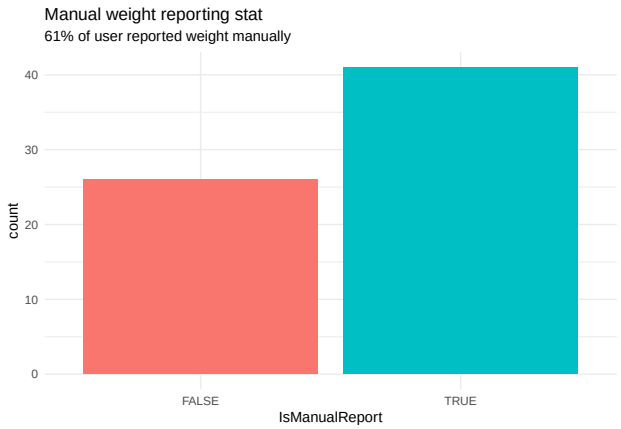


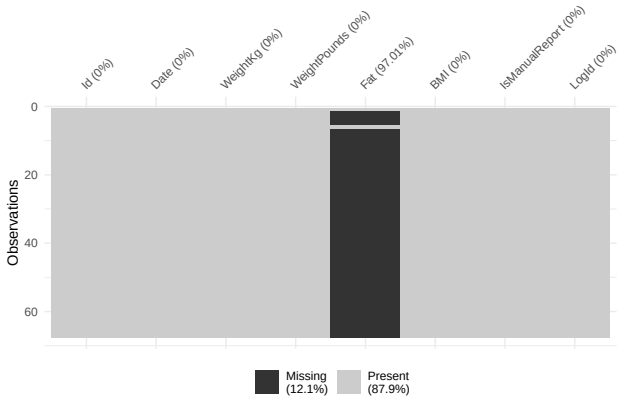
If we correlate the daily activity data and the sleep data we found surprisingly that there is a negative correlation between the number of workouts and the average sleeping time. This means that who workout more sleep less which is anti-intuitive.



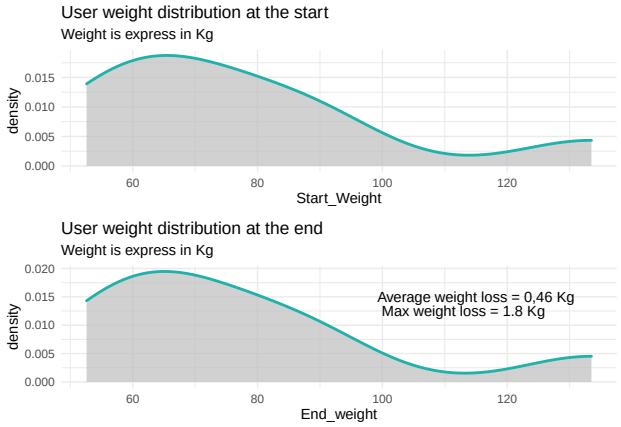
4.3 Weight loss anlysis

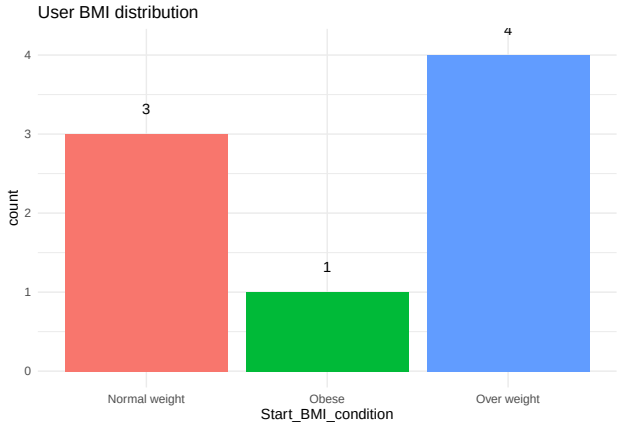
The weight loss data was the most controversial because was partially submitted by hand from the user. In fact almost all the Fat value was missing probably due to the absence of the right balance capable of estimate the fat presence inside the body.



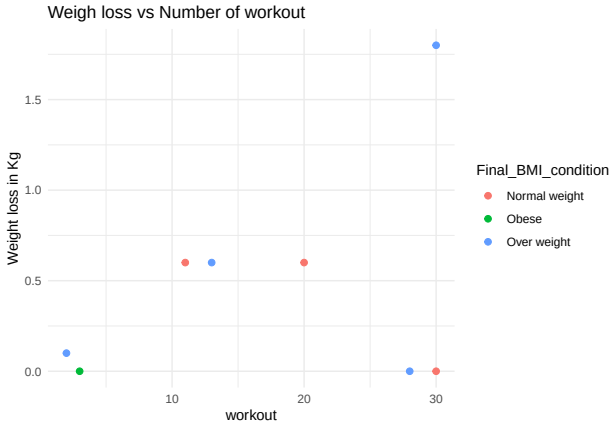


Observing how the weight of the users progress during the time period we found a very small difference because the majority of the users maintain the same weight with an average loss of only 0,46Kg (~1 pound).

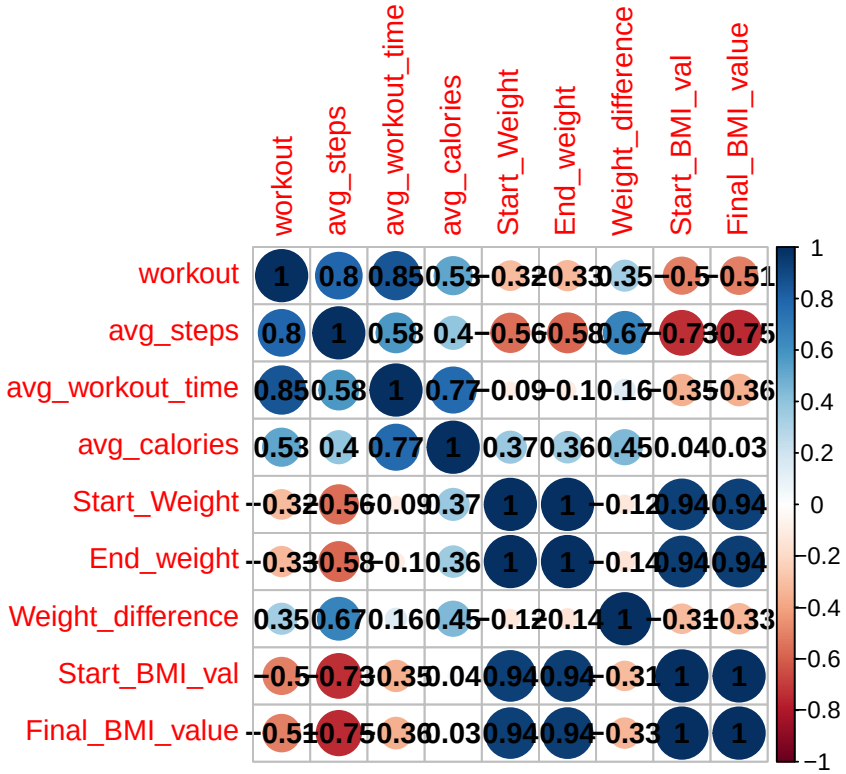




If we merged the data from the daily activity with the weigh loss data we discover something interesting: Some user even if the workout constantly everyday for a month, the weight loss was zero o very marginal. This could be due to a wrong nutrition or a specific training for building up muscles.



Proceeding with a correlation analysis, there is a very strong negative correlation between the average daily steps quantity and the BMI value which suggest, as you would expect, that walking is very good for health.



5.0 Recomandation

According to what we see in data my recommendations are:

- The data set provided was very limited in time frame and significant changes in fitness habits are difficult to manifest in this short time. A new analysis should be performed with a new data set to discover more information and confirm or dispute the findings of this work. According to the work **P.Lally et all 'How are habits formed: Modelling habit formation in the real worl'**, 66 repetitions was the average for activities performed daily becoming automatic. If we consider an average of 3 workouts per week, the ideal time to see possible changes or trend in user health is at least 22 weeks.
- The new data set should have information about user anagraphic. In this way more evident user cluster could be be identify an used for advertising.
- The favorite day for workout is Tuesday, this could be exploited by marketing.
- The weight loss data presented many missing fat values and 61% of user inserted data manually. This could be a market opportunity for a smart scale to be added inside the Bellabeat product list.
- Some users even if they train constantly for all the time period didn't loss weight. This could be due to a wrong nutrition or a specific training for building muscle. If further investigation confirm this, nutrition or training suggestion might be a key added value inside Bellabeat membership.